Contribution ID: **94** Type: **Presentation**

dCache: storage for XFEL scientific use-cases and beyond

Wednesday, 10 October 2018 14:30 (15 minutes)

The dCache project provides open-source storage software deployed internationally to satisfy ever more demanding scientific storage requirements. Its multifaceted approach provides an integrated way of supporting different use-cases with the same storage, from high throughput data ingest, through wide access and easy integration with existing systems. In supporting new communities, such as medical research, photon science/XFEL and microbiology, dCache is evolving to provide new features and access to new technologies.

Whatever the use case, for federated storage to work well some knowledge from each storage system must exist outside that system. This is needed to allow coordinated activity. To support such scenarios dCache provides a stream of internally generated events. In this approach the storage systems (rather than the clients) become the coordinating service, notifying interested parties of key events.

Storage events are also useful in other contexts: catalogues are notified whenever data is uploaded or delete, tape becomes more efficient because analysis can start immediately after the data is on disk, caches can be "smart" fetching new datasets pre-emptively and removing cached content when the source is deleted.

In this paper we will present work done at DESY in building a low-latency, compute cloud facility for various XFEL workflows. This was achieved by combining dCache storage events with various Open Source projects, such as Apache Kafka, Apache OpenWhisk and Kubernetes. The resulting "serverless" cloud service is similar to AWS Lambda or Google Cloud Functions. It allows the infrastructure to deploy additional resources automatically, seamlessly scaling to match the demand.

Type of abstract

Presentation

Summary

At DESY we are building building a
low-latency, compute cloud facility for various XFEL workflows. This was achieved by combining dCache storage events with various Open Source projects, such as Apache Kafka, Apache OpenWhisk and Kubernetes. The resulting "serverless" cloud service is similar to AWS Lambda or Google Cloud Functions. It allows the infrastructure to deploy additional resources automatically, seamlessly scaling to match the demand.

Primary authors: SCHLUENZEN, Frank (DESY); STAREK, Juergen (DESY); Dr SAHAKYAN, Marina (DESY); Mr SCHUH, Michael (DESY); Dr ADEYEMI, Olufemi (DESY); FUHRMANN, Patrick (DESY); MILLAR, Paul (DESY); YASAR, Sibel (DESY); Mr MKRTCHYAN, Tigran (DESY)

Presenters: Mr SCHUH, Michael (DESY); FUHRMANN, Patrick (DESY)

Session Classification: Data Management Services

Track Classification: Area 5. Digital Infrastructures for EOSC and/or EDI